Comparison of industrial hygienists' exposure evaluations for an epidemiologic study

by Patricia A Stewart, PhD,[1] Rafael Carel, MD,[2] Catherine Schairer, PhD,[1] Aaron Blair, PhD[1]

Stewart PA, Carel R, Schairer C, Blair A. Comparison of industrial hygienists' exposure evaluations for an epidemiologic study. Scand J Work Environ Health 2000:26(1):44—51.

Objectives A study was conducted to determine what level of information is required by industrial hygienists before they can develop exposure estimates comparable with those developed from a more in-depth evaluation.

Methods Three industrial hygienists evaluated formaldehyde exposures of 300 jobs selected from an earlier epidemiologic study. The jobs were evaluated over the following 6 cycles: (i) job title and industry; (ii) job title, industry, dates; (iii) job and department title and industry; (iv) cycle 3 information with dates; (v) cycle 3 information with a plant report; and (vi) job and department title, industry, dates, and the report. Each hygienist assigned jobs to 1 of 4 exposure categories, which were compared with the categories in the original epidemiologic study.

Results Overall, the mean differences between the hygienists' evaluations and the standard, although small, changed little over the cycles. The kappa statistic was poor to moderate for all the cycles, but the agreement was greater than expected due to chance. There was moderate improvement in overall agreement over the cycles using the weighted kappa statistic, but little improvement in the intraclass correlation coefficients of the hygienists' evaluations, which ranged from 0.4 to 0.5. Department information improved the agreement with the standard by 5—10%, but dates did not the improve agreement. There were some differences by type of plant, job function, exposure level, and date of the estimate. Using a hypothetical exposure-response scenario, this level of misclassification would have resulted in missing an association.

Conclusions Although there was slight improvement with increasing levels of information, these findings suggest that the subjective categorical assessment of exposures by industrial hygienists will not produce exposure estimates comparable to more in-depth evaluations of exposure.

Key terms exposure assessment, formaldehyde.

One of the most difficult and controversial components of occupational epidemiologic studies is the evaluation of retrospective exposures. This situation is probably due to the usual lack of sufficient measurement data and of recognized, established methods of estimation. As a result, the quality of the evaluation is heavily dependent on what measurement and other descriptive information about the work environment is available.

In the cohort study design, investigators often make site visits to the workplace, collecting historical documents on measurements and on production-related information, conducting interviews, and taking measurements.

This approach is very time-consuming (1, 2) and requires evaluation of large amounts of data. In addition, it generally must be done by an industrial hygienist or other professional of similar training and is, therefore, expensive. Thus it is important to determine whether procedures of less intensity provide a similar level of accuracy. This paper describes an exercise in which increasing amounts of information were provided to industrial hygienists estimating historical exposures to learn at what level of information they can develop estimates that are comparable with those developed using a more in-depth and intense effort.

[1] Environmental Epidemiology Program, National Cancer Institute, Bethesda, Maryland, United States.
[2] Department of Occupational Medicine, Division of Community Health, Faculty of Health Sciences, Ben Gurion University, Beer-Sheva, Israel.

Reprint requests to: Dr PA Stewart, Environmental Epidemiology Program, National Cancer Institute, 6120 Executive Blvd, Room 8102 MSC 7240, Bethesda, MD 20892—7240, United States. [Email: stewartt@epndce.nci.nih.gov]

Subjects and methods

In the original study, 2 industrial hygienists visited 10 companies that made formaldehyde and resins; resins, molding compounds, and molded plastic products (designated as resins and other products in this report); photographic film; plywood; and decorative laminates (1). They conducted walk-through surveys of the worksite, collected historical documents and results of formaldehyde measurements taken by the company, and conducted interviews with long-term workers. This process allowed them to obtain information on job descriptions, tasks, process flows, changes in the process or engineering controls, personal protective equipment, and other chemical exposures. In addition, airborne formaldehyde exposure measurements were taken. From this information they developed quantitative formaldehyde exposure estimates for every job/department/plant/year in the study. These estimates were then assigned to 5 8-hour time-weighted average (TWA_{8h}) exposure categories (1 = trace, 2 = 0.01—0.10 ppm, 3 = 0.11—0.50 ppm, 4 = 0.51—2.00 ppm, and 5 = >2.00 ppm).

Four industrial hygienists were hired to conduct the present study. All of them had several years of experience in assessing current exposures, but none had experience in assessing historical exposures. They had familiarity with 1 to 4 of the 6 industrial processes in the study. Prior to the start of this study, the industrial hygienists were given reports from the literature and permission to conduct a brief (<10 hour) review of these and any other data to which they had access. No restrictions were placed on the industrial hygienists as to what information they could review other than that they were not to review reports on these particular companies or on this particular study.

A random sample of 30 jobs from each plant was taken from the jobs in the original study (3) (table 1). Three industrial hygienists estimated the TWA_{8h} exposure levels of the 300 jobs by identifying an exposure category of 1 (the former categories of trace and <0.10 ppm) through 4 (the former categories of 2—5). The evaluations were made in the following 6 cycles of increasing amounts of data: (i) job title and type of industry, (ii) job title, type of industry and dates the job was held, (iii) job title, department title and type of industry, (iv) the same information as in cycle 3 with dates, (v) the same information as in cycle 3 with the addition of a brief report describing the processes at each plant, and (vi) job title, type of industry, department title, dates and the brief report. The reports, which varied between 4 and 12 double-spaced pages, had been developed for the original study, and they described the processes, jobs, and measurement data. Each cycle was started after the previous one had been completed and submitted to the study investigators, a process that prevented any modification of

the completed data set based on the new information. There was no feedback to the industrial hygienists as to their performance, and they developed the estimates independently of each other. The industrial hygienists also evaluated the level and frequency of peak exposures above the TWA_{8h} for each of the 6 cycles. A 4th industrial hygienist estimated exposures once for the same 300 jobs using only the information provided in cycle 6 to determine if any observed improvement was due to a learning curve.

To determine validity between the estimates developed by the industrial hygienists of this study with those of the original study (the standard), several agreement measurements were evaluated. To measure changes in both exact and approximate agreement, the mean and standard deviation of the differences were calculated between the standard exposure category and that assigned by the industrial hygienists (4, 5). Relative bias was calculated as the mean of the differences between the standard and the estimate divided by the mean of the standard values. Relative imprecision was the standard deviation of the differences divided by the mean of the standard values. The kappa statistic (k) and the weighted kappa statistic (k_w) were used to determine the proportion of exact and overall agreement, respectively, that could not be expected by chance alone (5). The kappa statistic has been interpreted as follows: >0.75 = excellent agreement,

Table 1. Some of the 300 jobs rated by the industrial hygienists.

Department	Job
Formaldehyde or resin plants	
Chemical laboratory	Chemist
Maintenance	Boilermaker
Polyol	Operator A
Bldg 77	Kettle operator
Purchasing	Purchasing agent
Molding compound	Production manager
54	Operator A
74	Filter operator
Resin and other products	
Mixing	Helper
Thermo plastics	Compex operator
Finishing	Buffer
Machine shop	Foreman
Molding	Helper
Material	Roll operator
Office	Timekeeper
Plywood	
Prefinishing	Press crew
Sawing	Bottom sander
Veneering	Taper machine lay up
Photographic film	
Chemicals	Supervisor
Coating	Coating cleaner
Film emulsion	Supervisor
Finishing	Slitter
Decorative laminates	
Mold	Tube rod
Press	Assembler
Product control	Press room planner

0.4—0.75 = fair to good agreement, and <0.4 = poor agreement (6). The weighted kappa was calculated by squaring the deviation of the pair of observations from exact agreement (ie, weighing the estimates closer to the true value more than those further away). To identify the areas of agreement and disagreement with respect to the level of exposure, conditional kappas were also calculated between the 3 industrial hygienists and the standard (7). The hypothesis of random agreement of the group of industrial hygienists relative to the standard was tested by calculating the G statistic (8). This statistic indicates whether the level of agreement exceeds that expected by chance. We are not aware of any interpretation of the magnitude of the G statistic, similar to that of the kappa. Reliability of the industrial hygienists among themselves (interrater agreement) was evaluated by the intraclass correlation coefficient (ICC) (9). It is interpreted like the kappa.

Analyses were performed by exposure level, type of plant (formaldehyde and resins, resins and other products, plywood, photographic film, and decorative laminates), the job function, exposure level, and date of the estimate. For the job function analysis, jobs were grouped into function (eg, maintenance or administration). Only job functions with more than 100 observations are presented. For the date of the estimate, 1970 was selected because of the likelihood of the increased awareness of occupational exposures (and therefore more effort to control exposures) that occurred around that time due to the promulgation of the Occupational Safety and Health Act.

Results

The mean differences between the category scores for the standard and for the industrial hygienists were small when compared with the mean of the standard values (1.83) (table 2). In the first cycle the overall difference

was -0.22, but this value decreased to -0.08 for the second cycle, and then there was very little change. The standard deviations in the estimates around the mean ranged from 0.8 to 1.0 across all the cycles, although it was lower in cycles 5 and 6 than in the others. These results represent a relative bias of less than 5% and an imprecision of about 40% after cycle 1. There was only a slight improvement in the exact agreement (k) between the estimates and the standard, going from 0.22 in the first cycle to 0.32 in the last cycle. The overall agreement (k_w), however, showed more improvement, increasing from 0.38 to 0.59. Agreement among the industrial hygienists, as measured by the intraclass correlation coefficient, changed little. The G statistic was significant for all the cycles and all the industrial hygienists, a finding indicating that agreement with the standard was not due to chance (not shown).

Industrial hygienist 1 tended to overestimate the exposures and did not improve as more information was obtained. Industrial hygienist 2 tended to underestimate the exposure estimates. This rater demonstrated little change in the exact agreement but improved somewhat in overall agreement. Industrial hygienist 3, who tended to overestimate the exposures, improved considerably after the first cycle both in terms of exact and overall agreement.

The amount of agreement for the level and frequency of peaks was similar to that of the TWA_{8h} estimates (not shown). There was slight improvement in cycles 5 and 6, similar to what was seen for the TWA_{8h}. No further analyses were therefore conducted.

The industrial hygienists assigned the same exposure category to 70—80% of the estimates from cycle to cycle. Of the remaining 20—30% they generally increased the level of their estimates when they had information on dates (the even numbered cycles) compared with the preceding cycle without dates (table 3). For example, going from cycle 1 to cycle 2, more estimates increased in level (N=185) than decreased (105). In each of the 3

Table 2. Agreement of the industrial hygienists with the standard and with each other, by cycle.[a] [k = kappa, k_w = weighted kappa, ICC = intraclass correlation coefficient, - = industrial hygienist on average underestimated the standard, + = overestimation; 0 (in parentheses) = no difference]

Cycle	All					Industrial hygienist 1[b]		Industrial hygienist 2[b]		Industrial hygienist 3[b]	
	Mean [c]	SD	k	k_w	ICC	k	k_w	k	k_w	k	k_w
1	-0.22	0.99	0.22	0.38	0.42	0.30(0)	0.42	0.34(-)	0.48	0.03(+)	0.25
2	-0.08	0.90	0.25	0.43	0.46	0.23(+)	0.38	0.28(0)	0.47	0.23(+)	0.43
3	0.04	0.94	0.22	0.37	0.48	0.24(0)	0.40	0.25(-)	0.43	0.19(+)	0.27
4	0.06	0.91	0.26	0.39	0.42	0.30(0)	0.37	0.26(-)	0.45	0.23(+)	0.35
5	-0.06	0.76	0.35	0.61	0.48	0.30(+)	0.55	0.42(0)	0.73	0.32(+)	0.55
6	-0.04	0.78	0.32	0.59	0.51	0.27(+)	0.48	0.37(0)	0.61	0.32(0)	0.69

[a] The G statistic exceeded statistical significance in all cases, indicating that the agreement was greater than expected due to chance.
[b] Number of estimates = 5400.
[c] Mean difference between the exposure category of the standard and that assigned by the industrial hygienists. The magnitude of the differences should be compared with the mean of the standard values (1.83).

cycles without dates the industrial hygienists tended to decrease the level of their estimates compared with the preceding cycle with dates.

The mean differences varied by type of plant over the cycles (table 4). The estimates for the resins and other products and plywood categories increased in agreement with the standard across the cycles. The mean difference went from -0.13 to 0.07 (relative bias = -4% to 2%) and -0.67 to -0.02 (relative bias = -28% to -1%) over the 6 cycles for these 2 plant types, respectively. The standard deviation also decreased over the 6 cycles for these 2 plants, falling from 1.23 to 0.76 and 0.90 to 0.49, respectively (relative imprecision decreased from 40% to 20% for both). This improvement was also seen for the 2 kappa statistics and the intraclass correlation coefficient. There were no other consistent patterns for the other types of plants. Surprisingly, the mean differences for the laminates and film plants increased over the cycles, although the other statistics generally showed some improvement. The G statistic was significantly different from chance for all cycles of the plant analyses except for cycle 1 for the resins and other products and for cycle 3 for film operation.

The level of agreement for the production jobs in each plant was similar to that observed for the type of plant, although there was a general, if inconsistent, pattern of slightly lower agreement for these jobs compared with that of the overall plant (table 5). (Jobs from the formaldehyde operations were not included because of small

Table 3. Direction of the change in the estimation level, by cycle.[a]

Cycle	Number of estimates		
	Increased	Decreased	Direction
1 to 2	185	105	+
2 to 3	178	200	-
3 to 4	106	85	+
4 to 5	154	240	-
5 to 6	157	132	+

[a] Nine hundred estimates were developed. The remaining estimates not indicated in the table did not change.

Table 4. Agreement of the industrial hygienists with the standard and with each other, by cycle and type of plant.[a] (k = kappa, k_w = weighted kappa, ICC = intraclass correlation coefficient)

Cycle	Resins or other products[b]					Plywood[c]					Formaldehyde or resin[b]					Laminates[c]					Film[c]				
	Mean[d]	SD	k	k_w	ICC	Mean[d]	SD	k	k_w	ICC	Mean[d]	SD	k	k_w	ICC	Mean[d]	SD	k	k_w	ICC	Mean[d]	SD	k	k_w	ICC
1	-0.13	1.23	0.05	0.15	0.36	-0.67	0.90	0.23	0.33	0.13	-0.34	0.82	0.34	0.47	0.52	-0.01	1.01	0.13	0.38	0.10	0.02	0.69	0.14	0.12	0.34
2	-0.12	1.14	0.13	0.25	0.39	-0.42	0.64	0.23	0.38	0.46	-0.20	0.78	0.35	0.48	0.49	0.27	0.88	0.13	0.32	0.22	0.65	0.62	0.18	0.15	0.46
3	0.22	1.23	0.09	0.12	0.44	-0.48	0.65	0.32	0.35	0.42	0.08	0.75	0.32	0.51	0.59	0.07	0.77	0.30	0.58	0.04	-0.08	0.78	0.03	0.04	0.38
4	0.26	1.22	0.12	0.12	0.36	-0.35	0.63	0.36	0.35	0.40	0.06	0.68	0.36	0.58	0.57	0.14	0.77	0.32	0.61	0.17	-0.07	0.70	0.08	0.19	0.37
5	0.07	0.83	0.31	0.65	0.51	-0.05	0.54	0.57	0.56	0.28	0.04	0.67	0.41	0.60	0.57	-0.10	0.77	0.38	0.63	0.35	-0.46	0.77	0.13	0.28	0.51
6	0.07	0.76	0.37	0.70	0.59	-0.02	0.49	0.51	0.59	0.28	0.19	0.69	0.33	0.51	0.58	-0.15	0.75	0.25	0.67	0.34	-0.61	0.86	0.12	0.29	0.48

[a] Only cycle 1 for the resins and other products plants and cycle 3 for the film plant were not statistically significant for the G statistic.
[b] Number of estimates = 1620.
[c] Number of estimates = 540.
[d] The mean differences should be compared with the following: resins and other products = 2.90, plywood = 2.37, formaldehyde or resin = 2.74, laminates = 2.57, and film = 2.53.

Table 5. Agreement of the industrial hygienists with the standard and with each other, by cycle and type of production job.[a] (k = kappa, k_w = weighted kappa, ICC = intraclass correlation coefficient)

Cycle	Resins and other products[b]					Plywood[c]					Laminates[d]					Film[e]				
	Mean[f]	SD	k	k_w	ICC	Mean[f]	SD	k	k_w	ICC	Mean[f]	SD	k	k_w	ICC	Mean[f]	SD	k	k_w	ICC
1	-0.58	0.97	0.20	0.21	0.21	-0.69	0.90	0.11	0.16	0.06	0.14	1.07	0.06	0.22	0.32	0.15	0.96	0.24	0.10	-0.01
2	-0.46	0.92	0.19	0.21	0.15	-0.49	0.63	0.08	0.33	0.21	0.51	0.78	0.08	0.24	0.36	0.19	0.86	0.23	0.10	0.13
3	-0.35	0.89	0.23	0.23	0.52	-0.64	0.74	0.07	0.02	0.13	0.14	0.75	0.31	0.56	0.24	0	1.10	0.09	0.01	0.02
4	-0.31	0.83	0.27	0.28	0.28	-0.42	0.69	0.06	0.03	-0.01	0.28	0.67	0.31	0.62	0.19	0	0.94	0.02	0.17	0.09
5	-0.03	0.77	0.34	0.49	0.37	-0.07	0.58	0.54	0.47	0.04	-0.13	0.77	0.36	0.56	0.29	-0.69	0.99	0.05	0.25	0.24
6	0.17	0.77	0.27	0.43	0.48	-0.04	0.42	0.59	0.64	0.71	-0.23	0.72	0.17	0.63	0.44	-0.72	0.94	0.07	0.32	0.51

[a] Only cycles 1-4 for the plywood plant and cycles 3-6 for the film plant were not statistically significant for the G statistic.
[b] Number of estimates = 1098. The number of estimates may not agree with those in the other tables due to a different distribution of unknown estimates identified by the industrial hygienists.
[c] Number of estimates = 270.
[d] Number of estimates = 324.
[e] Number of estimates = 306.
[f] The mean differences should be compared with the following: resins and other products = 3.02; plywood = 2.53, laminates = 3.00, and film = 3.06.

Table 6. Agreement of the industrial hygienists with the standard and with each other, by cycle and type of nonproduction job.[a] (k = kappa, k_w = weighted kappa, ICC = intraclass correlation coefficient)

Cycle	Administration[b]					Shipping, receiving, material handling[c]					Maintenance, labor[d]					Engineering and research[e]				
	Mean[f]	SD	k	k_w	ICC	Mean[f]	SD	k	k_w	ICC	Mean[f]	SD	k	k_w	ICC	Mean[f]	SD	k	k_w	ICC
1	-0.02	0.61	0.16	0.32	0.32	-0.53	0.84	0.10	0.40	0.21	-0.34	1.02	0.14	0.25	0.40	-0.03	1.09	0.14	0.06	0.36
2	0.13	0.59	0.16	0.09	0.25	-0.31	0.81	0.14	0.33	0.39	-0.33	0.88	0.18	0.37	0.44	-0.15	0.80	0.37	0.34	0.65
3	0.10	0.53	0.18	0.27	0.33	-0.19	0.88	0.17	0.29	0.09	0.10	0.85	0.16	0.35	0.24	-0.12	0.84	0.04	0.26	0.59
4	0.14	0.50	0.18	0.27	0.23	-0.17	80	0.13	0.39	0.10	0.06	0.93	0.17	0.31	0.33	-0.11	1.05	-0.03	-0.04	0.41
5	0.12	0.42	0.29	0.49	0.17	-0.17	0.73	0.12	0.53	0.42	-0.73	0.75	0.35	0.61	0.50	-0.11	0.83	0.24	0.43	0.38
6	0.05	0.42	0.34	0.59	0.32	-0.19	1.09	0.03	0.27	0.27	-0.11	0.74	0.25	0.60	0.44	-0.09	0.82	0.25	0.46	0.09

[a] Only cycles 1, 3, and 4 for engineering and research were not statistically significant for the G statistic.
[b] Number of estimates = 702. The number of estimates may not agree with those in the other tables due to a different distribution of unknown estimates identified by the industrial hygienists.
[c] Number of estimates = 360.
[d] Number of estimates = 774.
[e] Number of estimates = 252.
[f] The mean differences should be compared with the following: administration = 1.62; shipping or receiving or material handling = 2.20; maintenance, labor = 2.41; engineering and research = 2.14.

Table 7. Agreement of the industrial hygienists with the standard and with each other, by cycle and exposure level.[a] (k_c = conditional kappa)

Cycle	<0.1 ppm[b]			>0.1—0.5 ppm[c]			>0.5—2 ppm[d]			>2 ppm[e]		
	Mean[f]	SD	k_c	Mean[f]	SD	k_c	Mean[f]	SD	k_c	Mean[f]	SD	k_c
1	-0.63	0.81	0.32	-0.35	0.80	0.16	0.55	0.82	0.20	1.59	0.69	-0.01
2	-0.52	0.70	0.34	-0.15	0.69	0.23	0.70	0.74	0.18	1.51	0.56	-0.01
3	-0.48	0.69	0.34	0.07	0.74	0.15	0.71	0.72	0.29	2.06	0.47	0
4	-0.43	0.66	0.39	0.08	0.67	0.26	0.72	0.71	0.27	2.06	0.67	0
5	-0.42	0.62	0.44	-0.02	0.72	0.28	0.46	0.66	0.47	0.92	0.60	0.18
6	-0.41	0.63	0.42	0.06	0.69	0.27	0.46	0.72	0.36	0.94	0.67	0.23

[a] All the cells were statistically significant for the G statistic.
[b] Number of estimates = 1760. The number of estimates may not agree with those in the other tables due to a different distribution of unknown estimates identified by the industrial hygienists.
[c] Number of estimates = 1698.
[d] Number of estimates = 1326.
[e] Number of estimates = 204.
[f] The mean differences should be compared with the following: 1 (≤0.1 ppm), 2 (>0.1—0.5 ppm), 3 (>0.5—2 ppm), and 4 (>2 ppm).

Table 8. Agreement of the industrial hygienists with the standard and with each other, by cycle and date.[a] (k = kappa, k_w = weighted kappa, ICC = intraclass correlation coefficient)

Cycle	<1970[b]					≥1970[c]				
	Mean[d]	SD	k	k_w	ICC	Mean[d]	SD	k	k_w	ICC
1	0.19	1.02	0.24	0.36	0.45	-0.34	0.96	0.22	0.38	0.46
2	0.28	0.94	0.21	0.43	0.53	-0.19	0.86	0.25	0.41	0.51
3	0.34	1.01	0.22	0.34	0.58	-0.05	0.90	0.21	0.35	0.44
4	0.30	1.05	0.22	0.32	0.53	-0.01	0.85	0.26	0.38	0.44
5	0.22	0.82	0.33	0.59	0.50	-0.14	0.73	0.34	0.60	0.46
6	0.17	0.90	0.27	0.55	0.60	-0.10	0.73	0.33	0.59	0.44

[a] All the cells were statistically significant for the G statistic.
[b] Number of estimates = 1332. The number of estimates may not agree with those in the other tables due to a different distribution of unknown estimates identified by the industrial hygienists.
[c] Number of estimates = 4068.
[d] The mean differences should be compared with the following: <1970 = 3.43 and ≥1970 = 2.46.

numbers.) There were also no nonproduction jobs for which the agreement consistently improved (table 6). The improvement for nonproduction jobs was similar to that of production jobs. The G statistic was statistically significant for most of the production and nonproduction jobs.

Examination of the data by level of exposure found that most of the improvement was in the highest exposure category (table 7). The mean difference over the cycles changed in this category from 1.59 to 0.94 and the conditional k changed from -0.01 to 0.23 over the cycles. The other exposure categories also showed im-

provement, but it was not as great. The highest conditional kappas were found at the lowest exposure level, but these changed only slightly over the cycles. The estimates developed by the industrial hygienist for the >0.1—0.5 ppm category exhibited the lowest relative bias (<20%) of the exposure categories. The relative imprecision generally decreased as the exposures increased. The G statistic was statistically significant for all the estimates in this analysis, indicating that the agreement was not likely to be due to chance.

There was little change in the mean differences over the cycles for estimates before 1970, but the mean difference decreased from -0.34 to -0.10 after 1970 (table 8). There was little improvement in the exact agreement and only a moderate increase in overall agreement for estimates before 1970 or after 1970. The intraclass correlation coefficient increased from poor to moderate agreement before 1970 but changed little after 1970. The G statistic indicated that the agreement with the standard was not likely to have occurred by chance. The industrial hygienists tended to underestimate exposures before 1970 and overestimated them after 1970.

The agreement of the 4th industrial hygienist with the standard was lower than that of the other industrial hygienists in cycle 6 (not shown) and more closely matched the results of the other industrial hygienists for cycle 1 than cycle 6. The mean difference was 0.14, the standard deviation was 0.91 (relative bias of 8%, relative standard deviation of 50%), the kappa was 0.30, and the weighted kappa was 0.40.

The number of hours spent by each industrial hygienist was about the same for each cycle. Industrial hygienist 1 spent an average of 26.3 (standard deviation 2.6) hours on each cycle; industrial hygienist 2 spent 33.6 (2.8) hours/cycle; and industrial hygienist 3 spent 20.8 (6.9) hours/cycle.

Discussion

Careful exposure assessment of occupational epidemiologic studies is a tedious and complex process that can substantially add to the length and cost of a study. The assessment in the original study in this paper took almost 8 person-years. It is therefore important to know if the same quality of exposure assessment can be attained with less costly and time-consuming efforts. This study suggests they can not. Although the mean differences between the estimates of the industrial hygienists and those of the original study were small, they changed little over the cycles, they were imprecise, and they had a substantial effect on hypothetical relative risks. (See the following text.) For almost all of the subgroups evaluated, however, the estimates were in greater agreement with the standard than was expected by chance.

It is surprising that there was so little improvement over the cycles of increasing information. Exposure levels changed over time in most of the plants in this study, generally being higher in the past than at the close of the study in 1980. Having information on the date of the job (cycles 2, 4 and 6) did not appear, however, to change the validity of the estimates and therefore suggested that there were other factors that affected the assessments. The importance of date information in assessments of occupational exposures has, however, been seen in other studies (10). The inconsistency of our findings suggests more investigation is needed in this area.

At the end of the study the industrial hygienists reported that knowledge of the department name increased the confidence they had in their estimates, but in fact it had little impact on the agreement. It was expected that department name would make a difference because there were several departments in the study that made products not containing formaldehyde. Without a department title, the industrial hygienists were not able to distinguish between the production departments that handled or did not handle formaldehyde. The report provided some information on these departments, and therefore the mean difference for these nonformaldehyde departments decreased from -0.6 to 0.1 (not shown). Thus department title can have a substantial effect on the identification of whether exposure to a particular substance occurred. The lack of a similar level of effect among exposed jobs may have been because the department of many jobs may have been obvious (eg, chemist and boilermaker). The lack of an effect for these jobs may also have been because the department name of several jobs gave little information even with the job title (eg, operator A in department 54).

The industrial hygienists felt the information provided for their review prior to the start of the study was not helpful, even though it included most of the generally available literature and monitoring data on these industries from the Occupational Safety and Health Administration. This perception and the low agreement in cycle 1 suggests that reviewing the general literature may be of limited value in estimating exposure levels in a particular plant.

Surprisingly, however, even the plant reports had only a moderate effect on validity in most cases. There may be several reasons for this occurrence. These reports described the layout of the plant (ie, the type of operations in the various buildings), and the start-up date of the operations, process descriptions, primary job titles and exposure measurements where available. They were brief, however, and all the industrial hygienists expressed disappointment at their brevity. In addition, many of the job titles cited in the reports were not the same as those in the study. The titles in the study were those from the company personnel records. These often differed from the titles identified in the industrial hygiene records, the

source of the titles in the plant reports. In addition, the measurement data that were available were poorly documented, with the type of measurement (personal or area), duration, and representativeness of the sample usually missing from the report, their usefulness therefore being limited. Finally, the jobs were selected at random, but, as shown in table 1, they may have been particularly difficult to visualize, even with the reports.

The report did appear to result in a substantial improvement in the ability of the industrial hygienists to assess certain situations, in particular, the resins and other products and the plywood plants, and the highest exposed jobs. The improvement, however, may have been due to a learning curve. This possibility is supported by the 4th industrial hygienist's results. This industrial hygienist developed estimates only for cycle 6 (ie, with the plant reports). Agreement with the standard, however, was similar to that of the 3 industrial hygienists in the first cycle.

There were specific improvements with the extra information. Agreement for the plywood plant improved substantially with the availability of the report, primarily due to increased validity in the estimates of the production jobs. The report on this plant both provided measurement data and indicated that there were few changes in this plant over the years of operation. The industrial hygienists indicated that prior to reading the report, they had expected levels to be much higher (as reflected in the low kappas and negative mean differences in cycles 1—4). Once they knew that the levels were lower and that there was little change over time, they decreased the estimates, thereby increasing the exact agreement. Having measurement data has been found to improve exposure assessment (11, 12), probably because the data indicate the likely range of the exposures.

It was surprising to see the particularly poor agreement as more information became available on the film operations overall and the film production jobs in particular. Although none of the industrial hygienists had any experience with this type of operation, one would have thought that more information would have improved the agreement, not decreased it. The low agreement may, in part, be because these plants also had film operations that did not use formaldehyde, which could not be distinguished from the formaldehyde-using ones even with the department titles.

Administrative jobs showed one of the highest levels of agreement. This is not unexpected, as it can be assumed as a general rule that these jobs have no or low exposures. The fact that the analysis by exposure level did not show a similar pattern in the lower exposure group was probably due to the other job functions, such as utility, quality control, and others (not presented due to small numbers), but for which low agreement was found. Thus administrative jobs may be easier to assess

than other types of jobs. Interestingly, the kappas for administrative jobs were lower than many other job functions prior to the reports being obtained. An examination of the data, however, indicated that there were several cells with no data in the cross-tabulation of the standard values and the estimates; in this situation the kappa tends to be lower. In contrast to administrative jobs, jobs in maintenance, labor, and shipping, receiving, materials handling, engineering, and research were not generally evaluated well, probably because there can be such large variability, which makes such jobs more difficult to assess (2). The study results suggest that, when an assessment is done, more time should be spent on these types of jobs and little time should be spent collecting information on administrative jobs.

Few studies in the occupational literature have evaluated how well raters compare to a quantitative standard within the context of a cohort study or conditions similar to those of a cohort study (ie, information available on the specific worksite). Our finding that more information had little impact on exposure assessment has been found by others (13). These investigators found only a slight increase in intraclass correlation coefficients among experts assessing exposures to tasks over 3 cycles of increasingly more technical information (0.5—0.6 and 0.6—0.7 for dermal and respiratory exposures, respectively). They also found a decrease (from 0.4 to 0.2) in the kappa statistic over the 3 cycles when the experts' rankings were compared with a ranking of the tasks by measurement data. In another study the evaluations were made of current conditions. Industrial hygienists visited the plants and were given information about the plant and jobs (12). The standard was the mean of measurements taken at the same time. Agreement to quantitatively defined exposure categories ranged from 10% to 75% without measurement data and 20% to 100% with such data. Our results were slightly lower, but this is not surprising since the industrial hygienists in our study did not have any information available on the plants in the first 4 data cycles, had not visited the plants, and, in many cases, were estimating exposure levels occurring over 20 years ago.

Reliability among raters performing quantitative assessments with information on the specific worksite has also been examined rarely. A study of industrial hygienists evaluating exposures in a car assembly plant found levels of the intraclass correlation coefficient similar to what we found (-0.1—0.6), with most being around 0.2—0.4 (14). In that study, the industrial hygienists had information on the plants and substantially more measurement data than those in our study. In another study in which estimates of hours of exposure per year were estimated, the investigators found somewhat higher intraclass correlation coefficients (0.4—0.7) (15). Exposure hours, however, are likely to be more easily estimated

than quantitative exposure levels since hours is only one component that must be considered when exposure levels are assessed. De Cock et al (13) found intraclass correlation coefficients of 0.5 to 0.7 among experts ranking tasks.

It would be interesting to see how this degree of misclassification of the exposures would have affected the standard mortality ratios in the original study. This was not possible, however, because this methodological study involved only a small sample of the jobs in the study. To estimate the effect on health risks, data from the original study (3) were used to determine the number of nondiseased subjects, and the number of diseased was adjusted to create a hypothetical exposure-response relationship of 1.0, 0.8, 1.6, and 2.4. With the misclassification resulting from cycle 1, risks of 1.0, 1.0, 1.8, and 1.1 would be observed, whereas from cycle 6, risks of 1.0, 1.3, 1.4, and 1.0 would be observed. The observed higher risk in the 3rd category compared with the 4th is supported by the largest amount of misclassification in the >2 ppm category (table 7). The fact that the degree of misclassification in this study was similar to that seen in other studies suggests that we may have missed many exposure-disease associations due to our inability to estimate exposures well.

Whether there is any intermediate level of information that can be used to achieve closer levels of agreement to an in-depth investigation is not clear. Walkthough surveys could be an option. Use of other exposure information sources, such as more-detailed reports, monitoring data, or information on job tasks, is dependent on the plant being studied. For example, in this study, only limited monitoring had been done in these plants, and descriptions or detailed information on processes and changes in the worksite were not available in written form. Thus most of the information came from interviews. In contrast, for a study of acrylonitrile workers, 18 000 air measurements were available over a 10-year period (2). Written job and process descriptions and changes in the workplace were available in a few of the 8 plants in the study. Even so, most of the information was obtained through interviews. Therefore, unless much of the information had been developed prior to a study or assembled during a study by company personnel, there may be few other options.

There was little improvement in the agreement between the industrial hygienists' ratings and the reference data over the 6 cycles of information. The agreement was greater than what was expected by chance. When the data were examined by plant, job function, exposure level, and date, some improvements were discerned. The results of the study suggest that valid exposure estimates cannot be developed by industrial hygienists who assess jobs with only the title, type of industry, dates, and minimal information on the worksite being evaluated, compared

with evaluations conducted after a more in-depth study of the worksites.

References

1. Stewart PA, Blair A, Cubit DA, Bales RE, Kaplan SA, Ward J, et al. Estimating historical exposures to formaldehyde in a retrospective mortality study. Appl Ind Hyg 1986;1:34—41.
2. Stewart PA, Zaebst D, Zey JN, Herrick RF, Dosemeci M, Hornung R, et al. Exposure assessment for a study of workers exposed to acrylonitrile. Scand J Work Environ Health 1998; 24 suppl 2:42—53.
3. Blair A, Stewart P, O'Berg M, Gaffey W, Walrath J, Ward J, et al. Mortality among industrial workers exposed to formaldehyde. JNCI 1986;76:1071—84.
4. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986;1:307—10.
5. Maclure M, Willett WC. Misinterpretation and misuse of the kappa statistic. Am J Epidemiol 1987;126:161—9.
6. Siegel DG, Podgor MJ, Remaley NA. Acceptable values of kappa for comparison of two groups. Am J Epidemiol 1992; 135:571—8.
7. Bishop YMM, Fienberg SE, Holland PW. Measures of association and agreement. Cambridge (MA): The MIT Press, 1975:397—8.
8. Light RJ. Measures of response agreement for qualitative data: some generalizations and alternatives. Psychol Bull 1971;76:365—77.
9. Bartko JJ. The intraclass correlation coefficient as a measure of reliability. Psychol Rep 1996;19:3—11.
10. Dosemeci M, Cocco PL, Gomez M, Stewart PA, Heineman EF. Effects of three features of a job-exposure matrix (JEM) in risk estimates. Epidemiology 1994;5:124—7.
11. Hawkins NC, Evans JS. Subjective estimation of toluene exposures: a calibration study of industrial hygienists. Appl Ind Hyg J 1989;4:61—8.
12. Post W, Kromhout H, Heederik D, Noy D, Duijzentkunst RS. Semiquantitative estimates of exposure to methylene chloride and styrene: the influence of quantitative exposure data. Appl Occup Environ Hyg 1991;6:197—204.
13. De Cock J, Kromhout H, Heederik D, Burema J. Experts' subjective assessment of pesticide exposure in fruit growing. Scand J Work Environ Health 1996;22:425—32.
14. Macaluso M, Delzell E, Rose V, Perkins J, Oestenstad K. Inter-rater agreement in the assessment of solvent exposure at a car assembly plant. Am Ind Hyg Assoc J 1993;54:351—9.
15. Teschke K, Hertzman C, Dimich-Ward H, Ostry A, Blair J, Hershler R. A comparison of exposure estimates by worker raters and industrial hygienists. Scand J Work Environ Health 1989;15:424—9.